Modeling Medical Cost Trends for Advancing Age in the Long Run

Thomas E. Getzen*

Presented at the Living to 100 Symposium

Orlando, Fla.

January 8–10, 2014

Copyright 2014 by the Society of Actuaries.

All rights reserved by the Society of Actuaries. Permission is granted to make brief excerpts for a published review. Permission is also granted to make limited numbers of copies of items in this monograph for personal, internal, classroom or other instructional use, on condition that the foregoing copyright notice is used so as to give reasonable notice of the Society's copyright. This consent for free limited copying without prior consent of the Society does not extend to making copies for general distribution, for advertising or promotional purposes, for inclusion in new collective works or for resale.

* Professor of Risk, emeritus, Temple University. Executive Director, International Health Economics Association. contact: getzen@temple.edu

Modeling Medical Cost Trends for Advancing Age in the Long Run

Thomas E. Getzen
Professor of Risk, emeritus, Temple University
Executive Director, International Health Economics Association
contact: getzen@temple.edu

Abstract/Executive Summary
Medical costs are among the most significant factors in determining long-run fiscal requirements for the federal budget of the United States and for the individual household budgets of retirees. Rapid growth and high individual variance make projections of future expenditures in the 20- to 50-year range difficult and uncertain. This paper builds upon prior work to demonstrate that the apparent patter process changes systematically as the span and unit of observation changes— linear in the short run of months, morphing into a smoothed reflection of gross domestic project (GDP) growth over multiple years, and ultimately fitting a logistic growth curve over decades in the long run. A structural model of the health system as a function of decisions made at different levels of organization (individual, employer, group, city, nation, world) with effects that endure for differing amounts of time is constructed. The nature of medical transactions and the importance of budgetary boundaries are considered. Data from U.S. Health Expenditures 1929–2013 are used to makes estimates. Results show lags of three to six years, and suggest lower-frequency effects that last for decades. It appears the excess growth rates of medical costs during the 1970s and 1980s has moderated considerably, and is likely to continue to bend the medical cost curve downward. However, the budgetary impact of increasing longevity is apt to continue to increase over the next 50 years, by which time it is not implausible that the elderly could account for more than 50 percent of medical expenditures and require 13 percent of GDP funded through taxes or premiums. An extension of the model uses the current ratio of cost per person over/under age 65 to project that total expenditures on medical care for the elderly will approach $13 trillion by 2050. Questions are raised as to whether aging, medical costs, retirement and technological advances are best modeled as quasi-independent elements with causal effects, or as integral aspects of a joint-process characteristic of modern economies.

Modeling Medical Cost Trends for Advancing Age in the Long Run

Contents

Modeling Medical Cost Trends for Advancing Age in the Long Run

1 Introduction and Background

Health expenditures from the individual perspective look very different from the perspective of the nation as a whole. Patterns of medical cost growth look different when examined over a period of 10 months rather than 10 years, and different again when viewed over 10 decades. Articulating the links between perspectives and aligning the perceptions at different focal lengths is the purpose of this paper. The introductory section reviews prior studies showing that the medical system is slow to adjust and that insurance serves to reduce income and budget constraints for individuals but not the nation. Section 2 presents a formalized model of decisions and spending for the health care system, and discusses the limitations on empirical analysis of health care spending over time. Section 3 reviews results for three spans of observation: months, years and decades. It also addresses some problems of variable identification, with specific application to population aging and the aggregate cost of medical care for those age 65+. Section 4 extends the model to estimate the aggregate cost of care to those age 65+, demonstrating how sensitive expenditures are to the allocation of resources between old and young (over/under age 65). Section 5 explores the sustainability of current trends, the boundaries between long-term care, retirement and medical expenditures, and concludes by suggesting that rising longevity and medical costs are best viewed as aspects of a process of economic and human development transforming the 20th and 21st centuries rather than as isolated phenomena.

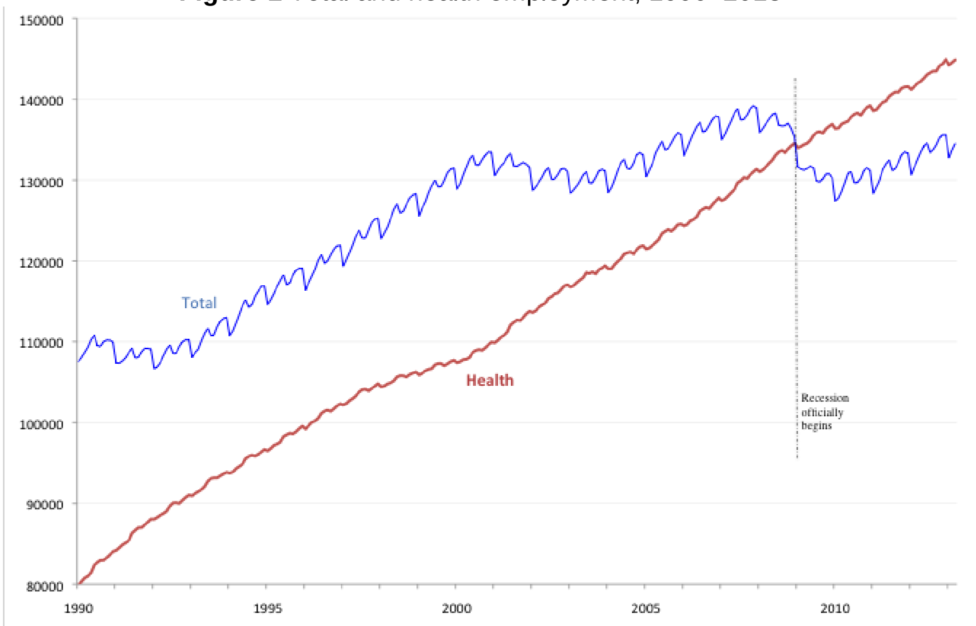
1.1 Medical Care is Slow to Adjust

The health care system is complex and slow to adjust. Coordinating thousands of patients, doctors, hospitals and payers is not just difficult, it is time consuming. This makes it almost impossible to respond immediately to a sudden change in the environment (such as a recession). A major change in the infrastructure of medical organization and payments will tend to persist, perhaps for decades, just because it is so difficult to achieve the public and political consensus required to bring about restructuring. For 50 years, the U.S. Congress has been tinkering with Medicare, chipping away at the foundations but unable to replace even those features that have become obviously outmoded, such as the Part A (inpatient)/Part B (outpatient) distinction.

The sluggish response of health expenditures to changes in the macro-economic environment can be graphically illustrated through employment and spending.

Figure 1 shows total employment and health employment in the United States and health employment from 1990 to 2013. It is evident the growth path of health employment is much smoother. A large dip due to the recession in 2008–10 is very evident for total employment, but is essentially invisible in the steady upward march of health employment. The saw-tooth seasonal variation is clear in the total series but reduced to a minor wiggle in the health series.

Figure 1 Total and health employment, 1990–2013



Source: U.S. Bureau of Labor Statistics at www.BLS.gov/data

Figure 2a shows the effect on national health expenditures of the major recession that struck Finland in 1991. The trough in health expenditures does not arrive until two years later. **Figure 2b** presents the same Finnish data but in the form of annual growth rates. Without taking account of the lag, the correlation between gross domestic product (GDP) and health spending is not apparent. However, a transformation of the income data using a three-year moving average of income lagged by two years as shown in **figure 2c** makes the correlation visible. The relevant variable is not the growth in this year's income, but the growth two, three and four years prior. There is an even stronger lag pattern in the response of spending to macro-economic changes for the United States (see section 3).

The complex and cumbersome systems by which medical care is financed are one cause of delays between changes in GDP and changes in health expenditures. Government and private insurance companies take time to respond, sometimes a great deal of time. Premiums and payment schedules are set a year or more in advance. Insurers have more incentives to wait and watch than to pay quickly. Third-party payment systems not only create delays, they change the relationship between buyer and seller, radically altering the market and the meaning of the word spending.

Figure 2a Effects of Finnish recession, 1980–2000

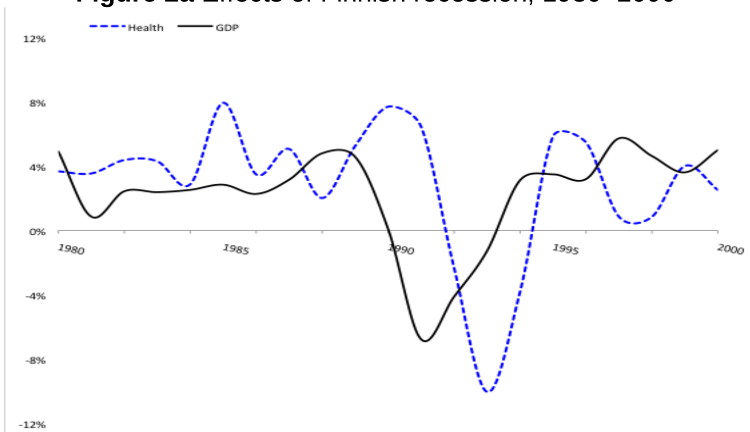


Figure 2b Current growth

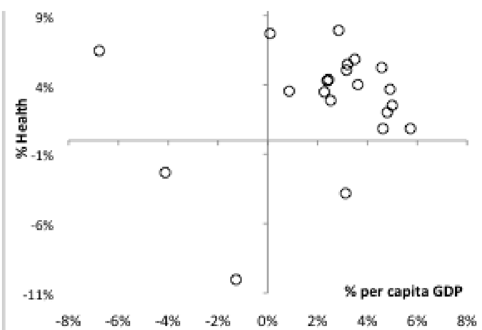
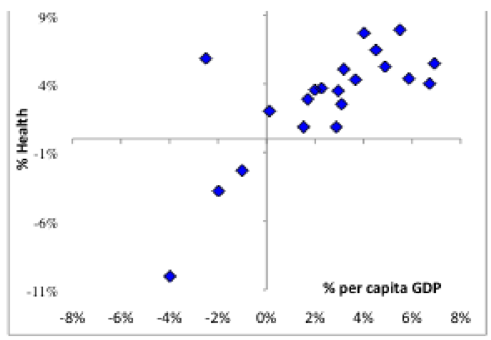


Figure 2c Lagged growth



Source: OECD Health Data at www.OECD.org/health

1.2 Boundaries and Budget Constraints

Insurance is necessary for the development and continued existence of a modern health care system. Pooled financing mechanisms capable of funding billions of dollars in medical education, research, hospitals, pharmaceuticals, nursing homes and diagnostic laboratories are needed. Without insurance, medical expenditures could not have grown to reach one-tenth of GDP, and certainly not continue growing as they have. Insurance can be mostly public and tax supported, private and employer supported, or a mixture of both as now prevails in the United States. Whatever the financing base, insurance is necessary to gather the funds of millions of individuals so that together they can obtain the benefits of a trillion dollar technologically and organizationally sophisticated medical care system. Such a complex professional, financial, scientific and organizational system takes time to adjust, and political consensus is required to undertake major structural changes.

Banking and other financial mechanisms allow consumers to optimize by shifting consumption between periods through savings and debt. Insurance allows consumers to optimize by shifting costs across members of the group. Before the advent of formal financial institutions, families and other social structures carried out these processes, expanding the decision set to allow shifting across time or across people. Insurance relaxes the individual and family budget constraint as costs are pooled over large numbers. If the insurance is paid for by the nation as a whole, then the actual budget constraint is national, not group or individual. The allocation of treatments is made by physicians based on each patient's illness and potential for benefit, not on that patient's income or budget. The decisions that determine the quantity of resources available and the quality of physician education are made at a higher level (hospital, employer, state, nation).

"Spending" is a decision allocating resources (money, time, water, space, broadcast spectrum) subject to a budget constraint. It is not an event, like mortality or morbidity. It is a decision, often a complex and collective decision, which is inherently social. More precisely, spending occurs with a transaction, often between two parties, but sometimes with multiple parties up to and including the whole of society. The simplest of markets are made up of "spot transactions," an exchange of a specified good for a specified price taking place at a single point in time and space. Many transactions, however, are complex exchanges involving multiple parties and lasting a long time, with contractual terms that are largely implicit rather than fully spelled out (rentals, hiring, partnerships, marriage, citizenship). Medical care is among the more complex of contracts. It is financed through third-party insurance, highly regulated, fraught with emotions and sacred if not religious overtones, and subject to an unspecified duty to "care" (Arrow 1983).

There is a subtle but significant difference between the amount spent on a patient and spending. The treatment is for an individual illness, but the payments and spending decisions are for physicians, hospitals, groups or systems and made by the family, the employer, the insurer or the government. The allocation of resources across patients is primarily determined by personal health status and professional judgments of physicians, while the total amount of resources available to allocate is determined at a larger level. Since most of the more localized levels of spending are regulated or subsidized, the inclusive total budget is determined nationally (Getzen 2000b). At the largest scale, medical science is created as a public good and funded by all the societies that carry out research and practice medicine around the globe.

Nations, unlike patients, face hard budget constraints. They must use taxes or premiums to pay for all of the medical care consumed within their boundaries. International financial markets may allow shifting of payment a few years forward or back, and international aid can be a significant source of medical funds for some poor nations, but in general the buck stops at the national boundary. Moving across a border from Kansas City, Mo., to Kansas City, Kan., or from Philadelphia to Camden, N.J., may have some effect on medical costs as the state line is crossed. Crossing over an international boundary from El Paso, Texas, to Juarez, Mexico, or from Detroit, Michigan to Windsor, Ontario, is literally a separate country in terms of health care spending. With a hard budget constraint in place, the amount of money available (per capita income) becomes the dominant determinant of total health spending (see **figures 3a and b**). That is why there is such a strong linear relationship between income per capita and medical spending per capita by nations, but not by individuals, families, cities, areas or regions.

Figure 3a Per capita GDP and national health spending across 20 countries, 2010

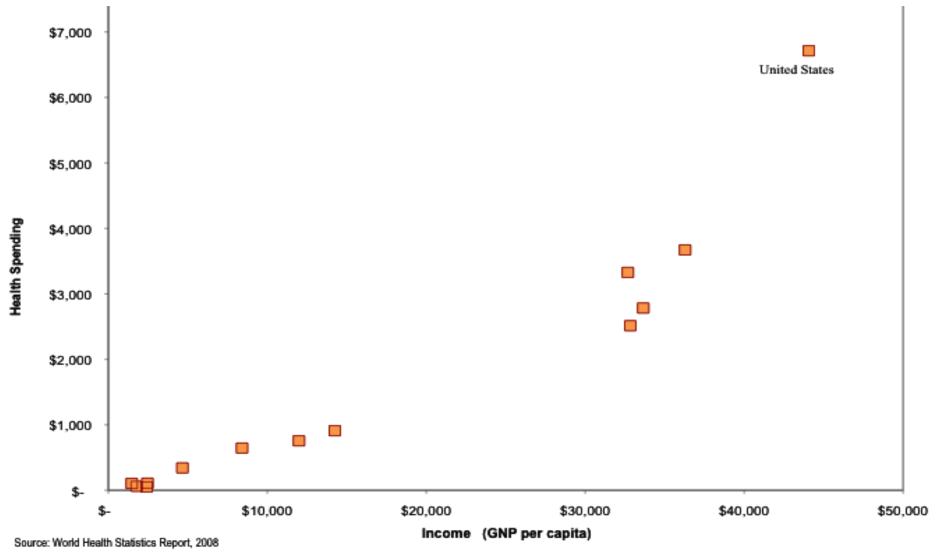
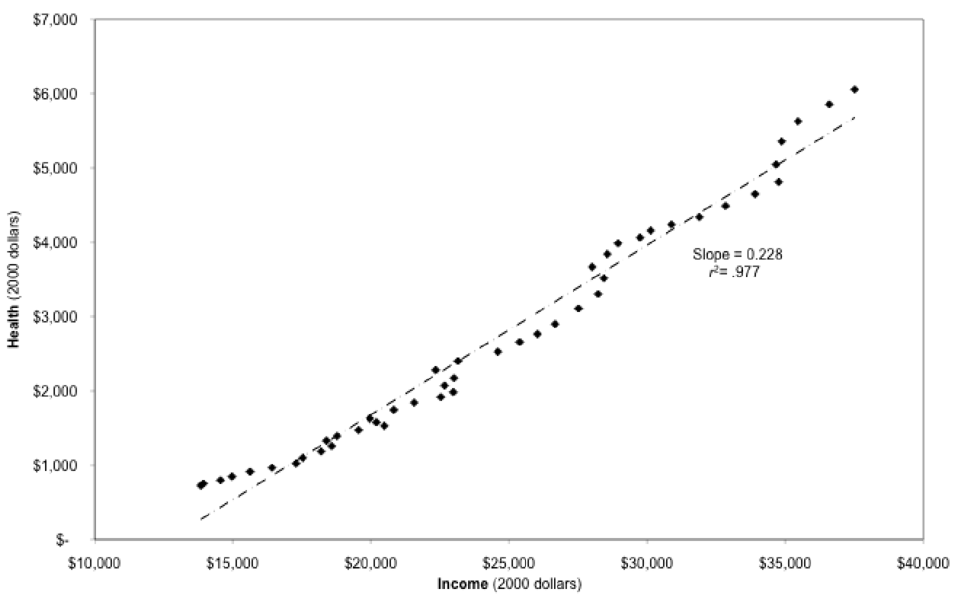


Figure 3b Per capita GDP and national health spending across the U.S., 1960–2005



Source: CMS Office of the Actuary at www.cms.gov/

2. Methodology

In this section, a model of medical spending is constructed in terms of decisions made at different times and at different levels of organization within the health system. Temporally, the units range from months to decades or even centuries. The levels range from individual patients to physicians, hospitals, regions, cities, nations and ultimately the world. Attention is paid to the boundaries at which budget constraints can be imposed. The growth of spending is related to a standard macro-economic decomposition in terms of population, prices and income per capita. The section ends by examining the implications of measuring and observing only reduced forms of the model.

2.1 Decisions and Transactions: Modeling Health Spending

The model described here is based upon that in Friedman (1957) with three modifications. The first modification is that Friedman focuses on a two-part allocation between spending on all other goods and services, whereas here the focus is on a two-part allocation between spending on all other goods and services and medical care. The second modification is the introduction of a layer of unobservable variables regarding decisions and organizational structure for expository purposes that do not affect the empirical analysis but aid in the explanation of the process and give some substance to the observed temporal lags. The third modification, which is significant, is that spending by individuals does not simply aggregate additively to equal national spending, hence the boundaries of the budget constraint must be more fully specified. The "decision-maker" could be a person, an employee group, a community or taxpayers and the nation as a whole. A formal analogy would be to extend the basic Friedman model to include pension and disability insurance payments. Pension and disability insurance, however, pay for only a small fraction of most ordinary consumption. Group insurance and taxes pay for more than three-fourths of medical care payment and hence affect the decision process much more heavily.

A skeletal description of the complex medical care spending process can be modeled as follows. Let:

$_i$ = individuals	$_N$ = nations
y = individual income	Y = national income
ψ = individual decisions	Ψ = national decision
h = individual medical spending	H = national medical spending

M = medical system

δ = tax rate

$_t$ = time (in years)

Z = exogenous parameters

f = fraction of the population age 65

Although individuals and nations must make expenditures from current income (supplemented by savings or debt), those decisions are usually based on their expectations of long-run permanent income, $\cancel{y}_t$ for individuals and $\cancel{Y}_t$ for the nation as a whole.

Expected permanent income: $\cancel{y}_t = f(y_t, y_{t-1}, y_{t-2}, \ldots y_{t-n}, z)$

$$\cancel{Y}_t = f(Y_t, Y_{t-1}, Y_{t-2}, \ldots Y_{t-n}, Z)$$

The nation spends from its budget, which is a fraction of total current income determined by the tax rate δ, and also has some ability to control individual spending through its management of the economy and rule-making powers (i.e., copayments, mandates, inflation). Individual citizens have some power over how the national budget is spent, but only through the political process.

National budget: $\delta Y_t = \delta \Sigma y_{it}$

The medical care system (M) at a point in time (t) is determined by the individual and national decisions made in prior years, and also by some exogenous parameters (Z) such as technology, biology, historical accidents and trade. Decisions are made atomistically by individuals (ψ_i) and politically by nations (Ψ_N) in each year. An individual making a decision at a point in time ψ_{it} does so based on the medical care system as it then exists (M_t), their estimated permanent income at that time ($\cancel{y}_{it}$), the decisions they have made in prior years, and other factors such as tastes, preferences and health status. National decisions are made each year (Ψ_{Nt}) and the budget (i.e., current income) is allocated with the goal of optimizing current and future states of the medical care system balanced against other goals and objectives (defense, education, debt, etc.). Similar to individuals, the nation must take the current overall state of the medical system as a given, but they also are attempting to optimize based on nation's estimate of how those decisions will affect the system in future years $\cancel{M} = (\cancel{M}_{t+1}, \cancel{M}_{t+2}, \ldots \cancel{M}_{t+n})$.

$$h_{it} = f(\psi_{it}, \{1-\delta\}y_t, M_t, \cancel{y}_t, z)$$

$$M_t = f(M_{t-1}, \Psi_t, \delta Y_t, \cancel{Y}_t, Z)$$

$$H_t = f(\delta Y_t, \Psi_t, M_t, \cancel{Y}_t, \cancel{M}, Z) + \sum h_{it}$$

Individual health spending depends on a set of individual decisions, disposable income, the state of medical organization/technology and other factors (including health status). The current state of medicine depends of the state of medical organization/technology inherited from the prior year, supplemented by some new decisions and financed with current (and future) national budgets. The nation's health spending is determined by its current and permanent income, the current state of medicine, and the expected effect of the decisions taken on future states of medical care, with a stub added to account for individual health spending.

These equations are largely recursive and cannot be solved—medical spending depends on a medical care system that is a complex function of prior and expected future states, which are themselves determined by decisions affecting organization and technology that are only intermittently and incompletely observable, and virtually never quantifiable. What the equations can do is isolate those observable variables and provide some insights into a temporal structure that must then be estimated.

2.2 Decomposition: Inflation, Population, GDP Plus Excess

Macro-economic rates of growth are usually decomposed into real and inflation components. For health expenditures, decomposition can usefully be taken one step further by defining the difference in rates of growth for health expenditures and GDP as "excess."

Let:

Y = nominal GDP	H = nominal health spending
g = growth in nominal GDP	h = growth in nominal health
r = growth in real GDP	x = "excess" growth (h – g)
d = inflation rate (GDP deflator)	S = health share of GDP (H/Y)
	s = growth rate of health share

The rate of increase in health spending is decomposed as *growth in health* = *inflation* x *real GDP growth* x *excess* and with compounding:

$$h = (1 + d)*(1 + r)(1 + x) - 1$$

And then *s*, the growth rate in the share of GDP, is readily seen as simply equal to *x*, the excess growth rate in health spending.

$$s = [H^*(1 + d)^*(1 + r)(1 + x)/Y^*(1 + d)^*(1 + r)]/(H/Y) - 1 = x$$

Graphs of shares are often more intuitive and quickly grasped than plots of nominal spending or "excess growth rate" since they are comparable across time periods, regions, countries or groups without regard to population, price levels, exchange rates or other factors that would require adjustment. Share estimates tend to be more reliable and available than spending estimates for historical periods and organizational subgroups, and illustrate:

- Spending relative to total resources available (budget constraint) at each point in time
- Curve is flat and horizontal where growth is equal to GDP + 0 percent
- Slope of the curve (percent growth rate in share) equals the rate of excess growth
- Maximum rate of excess growth as indicated by the inflexion point of the curve

2.3 Aggregation, Reduced Forms and Empirical Observations

The complexity of a health care system with decisions made at multiple levels having effects lasting for varying lengths of time presents daunting econometric challenges. The equations above can provide a conceptual map, but they do not easily yield a function form that can be estimated. Although in principle individual decisions could be observed, the use of third-party insurance financing means that most spending at the individual level does not come from disposable personal income, that personal (out-of-pocket) spending does not aggregate up to national spending. Instead, the share of health spending coming from personal budgets (disposable income), $\sum h_{it}/H_t$, is 15 percent or less for the United States and most other countries with advanced medical systems, with a significant portion of that influenced or compelled by third-party rules (co-payments, deductibles, mandates) rather than independent purchasing decisions made by individuals.

If a personal budget constraint does not apply, then the cost of medical treatment is not "spending" from an individual point of view. This consideration lies behind the first major simplification aimed at making the model applicable to empirical estimation. Since individual, group, city and regional decisions are not (or only partially) subject to a budget constraints, they do not constitute spending in the ordinary economic sense and so are all subsumed into the national level where budget constraints are binding.[1] Concerns about cost shifting, cost allocation, cross-subsidy or price transparency become irrelevant from a national perspective, except insofar as they serve to obfuscate budget realities from the public and patients.

Focusing on a national level that includes all geographic and organizational subunits simplifies and sharpens the temporal perspective as well. Although a patient will get sick and be admitted on a specific day, the hospital reimbursement implicitly and often explicitly covers an entire year—making smaller units of observation superfluous. The weekly rise and fall of patient flow does not need to be matched by a corresponding cycle in receipts. Even when reimbursement is tied to specific bills, the clearing of claims typically takes six to 18 months—a sufficient time that an insurers' investment earnings are routinely larger than any underwriting gains and losses.

Estimates of permanent income for individuals are intimately tied to a lifetime trajectory of earnings and the problem of retirement. This requires foresight that may stretch 50 years into the future, and for which mortality is a risk that can only be partially offset. Expected permanent

[1] This point is extensively examined along with the supporting empirical evidence in Getzen (2000a, 2006). The constituent decision-making cannot be divided into a set of parts and added up. A more basic explanation of the atomistic fallacy of generalizing from the individual to the group (e.g., since this person's costs rise a lot as they get older, then societies' average cost must rise as they age) and the wider application to studies of national health expenditures is provided in Getzen (2013a, 293–99).

income for the nation, in contrast, is usually a linear projection with steady annual average gains. Mortality and retirements are regular, continuous and relatively predictable over a population of millions. Recessions, however irregular, mostly create temporary fluctuations rather than the long-run disruption that unemployment often imposes on individual earnings. Aging is a significant and ultimately intractable problem for the individual, while population aging is a manageable problem and has had a much more modest impact on national health care cost trends than initially thought (see section 3).

The state of the medical care system (M_t) is a convenient construct but not observable or quantifiable. Decisions (Ψ_t) are only occasionally visible and have effects that may only become apparent with the passage of time. The two variables within the model that are observable and measurable are income and health spending. Everything else resides within a black box. Thus the only reduced form that can actually be estimated is:

$$H_t = f(Y_t, Y_{t-1}, Y_{t-2}, \ldots Y_{t-n}, Z)$$

Even though most of the internal structure is unobservable, the model does suggest why the health care system has such a long memory, and where it resides: in the structure of the medical care system itself. Major decisions[2] take a long time to craft, involve multiple and often competing coalitions, and affect the entire nation for decades or longer. Medicare was legislated 50 years ago as part of the Social Security Act but still exerts influence and imposes an obsolete financial distinction (Part A and Part B). The Graduate Medical Education National Advisory Committee

[2] Examples include the National Institutes of Health (1887 to 1950), physician licensure (1910 to 1935), laws such as Hospital Survey and Construction Act (Hill-Burton Act) of 1946, Social Security Act of 1965, Employee Retirement Income Security Act (ERISA) of 1974, Tax Equity and Fiscal Responsibility Act (TEFRA) of 1982 and the Patient Protection and Affordable Care Act (PPACA) of 2010.

(GMENAC) Report and subsequent medical workforce regulations from the 1980s still affect labor supply. The fiscal impact of any particular political decision will eventually decline, but a pattern of delays or lingering effects is itself evidence that the change was likely to be structural. Empirical estimates of such delays and interactions may eventually provide some insights into the workings of the black box.

3 Results: Matching Temporal Span and Units of Observation

Most of the empirical results illustrated and referred to below have been separately published and commented on by a number of authors. The task of this study is to bring the pieces together, make comparisons and demonstrate the relationship between micro and macro results. In particular, to illustrate how the relevant factors change as the scale of observation changes, and to make explicit the mechanisms that link short-, medium- and long-term results.

Health spending growth is best examined at three different time scales. From month-to-month, growth appears to be a straight-line process with steady accumulation. Over a span of several years or decades, a smoothed fluctuation reflecting the ups-and-downs of the business cycle becomes apparent. Examining the rate of excess growth over the last 50 years makes it evident that a longer-term process is occurring, but the shape of the curve only becomes obvious when the span of observation is a century or more.

As the unit and span of observation changes, different phenomena are observed. Temperatures measured hourly can show the passage of a cold front, while temperatures by decade or century are better for seeing the advent of global warming or a new ice age. It is obvious that minutes are too short, and millennia too long, to observe the effect of a recession on health spending. The optimum temporal unit of observation must obviously fall somewhere in between.

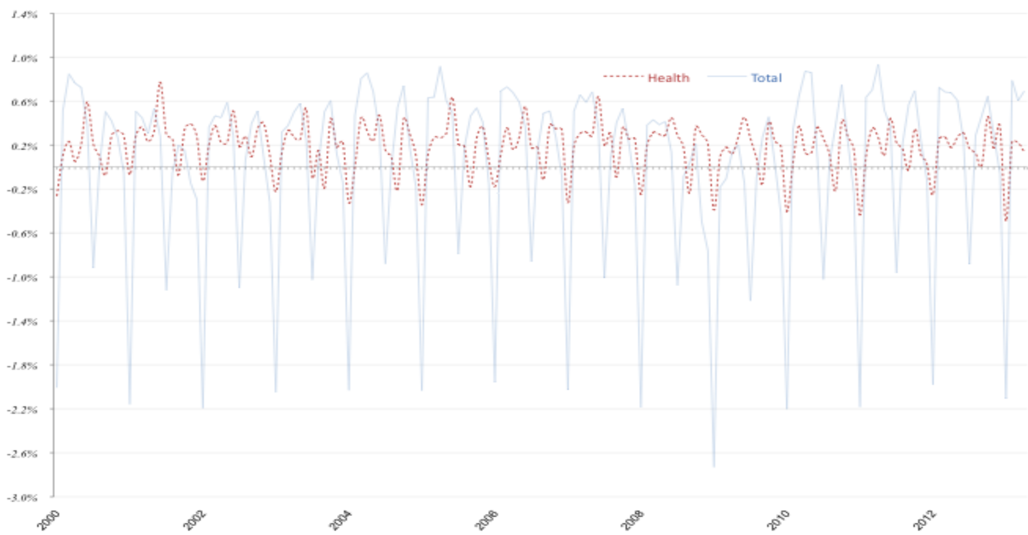
Although imprecise, a rule of thumb is to consider a span of observation within two orders of magnitude (×10 to ×100) of the expected length (or half-life) of the phenomenon, measured within one order of magnitude ($1/10^{th}$ to 1×). Recessions, as measured by the National Bureau of Economic Research (NBER), have lasted an average of 12 to 18 months. This approximate guide would suggest a span of observation not less than 10 or more than 150 years, with units not shorter than months or longer than years. In practice, NBER provides quarterly estimates over the last 150 years separated into spans of about 50 years.

3.1 Daily and Monthly Data

Although GDP data are available only quarterly and national health expenditure (NHE) data annually, the growth in employment, a primary element of economic activity, is available on a monthly basis for both total and health care employment, and can be used as a proxy for high-frequency changes. The current Bureau of Labor Statistics (BLS) health series begins in 1990 (the overlapping prior 1958–2002 series uses a slightly different definition of health care).

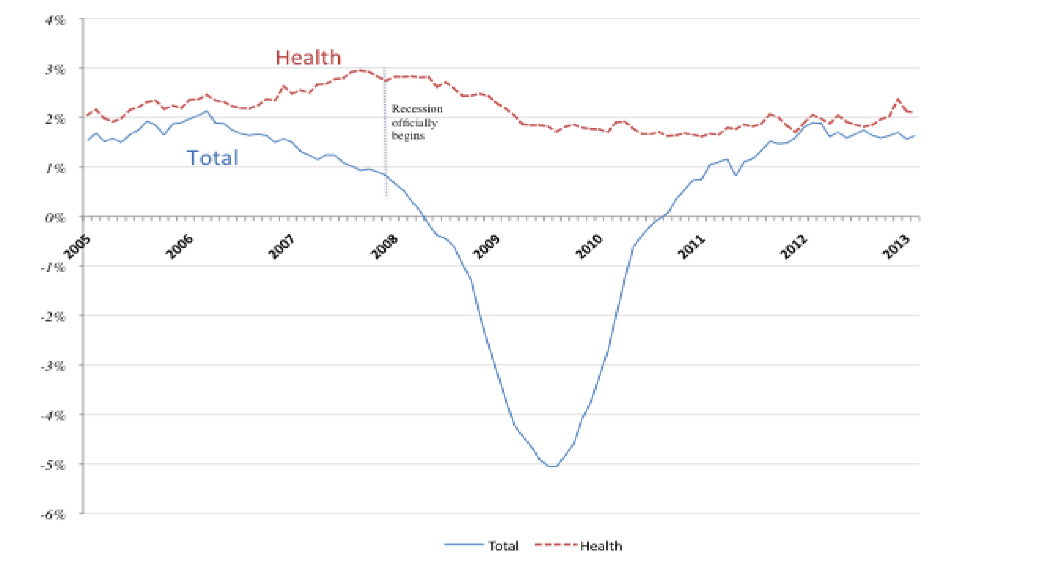
Monthly rates of change for both total and health care employment from January 2000 to April 2013 are graphed in **figure 4a**. It is evident that health series has smaller variance (.22 vs. .82) and less pronounced seasonality. Although perhaps not so obvious, average employment growth for health care is more than eight times as large (0.187 percent vs. 0.022 percent). The major features become much clearer in **figure 4b,** which smoothes and removes seasonality by a plotting a moving average of annualized growth rates over the preceding 12 months. The recessions of 2001 and 2007–09 are seen to have large effects on total jobs, driving the economy down into losses after some lag, but to have only slight and delayed effects on health care job growth, which consistently remains positive. The cumulative health care employment effect, a delayed 1 percent reduction over 10 to 36 months, is small and spread out enough that it might not be highlighted in a cursory analysis, or disregarded as not being statistically significant. Finding the delayed, smoothed correlation with total employment would be even more difficult, given the much larger variance in that series and the need to account for lags.

Figure 4a Monthly rates of growth in total and health employment, 2000–13



Source: U.S, Bureau of Labor Statistics.

Figure 4b Annual percent change in employment, 2005–13



Source: U.S, Bureau of Labor Statistics.

With a phenomenon (recession) that usually lasts 12 months or more, quarterly data are sufficient and the extra "data" in monthly differences is mostly noise. Since the health sector response takes even longer (two to six years), the span of observation required is longer, and the use of more frequent monthly observations even less helpful. Some employment data are available weekly, and data on spending could potentially be collected on a daily or even hourly basis. Such high frequency data will add noise but would not yield extra insight into low-frequency growth cycles or a phenomenon occurring over years or decades. What such flood of measurements could do is overwhelm analysis and generate lots of coincidences and chance hypotheses, much the way play-by-play commentary on a ball game generates endless speculations about the outcome or similarly intent watching of weather radar or financial broadcasts can lead to countless conjectures about the amount of rainfall or stock prices for the coming year.

3.2 Business Cycles and Variable Lags

If a researcher looked only at the contemporaneous correlation, they might conclude there was no relation between the health spending and the business cycle. The lag response smoothes the curve and shifts it in time so that more work is required to reveal the relationship. The effect of income growth on health spending is substantial but distributed over time.

The lag between the 1994 recession in Finland and the corresponding decline in health expenditure growth shown in **figure 2** was two to three years. The structure of the lags seems to vary somewhat between countries, with the United States showing a longer than average lag (peak response after about four years). Even within a specific country, the lag structure appears to vary across types of spending, with construction and inpatient services slower than average, while dentistry and out-of-pocket patient payments respond most rapidly (**table 1**).

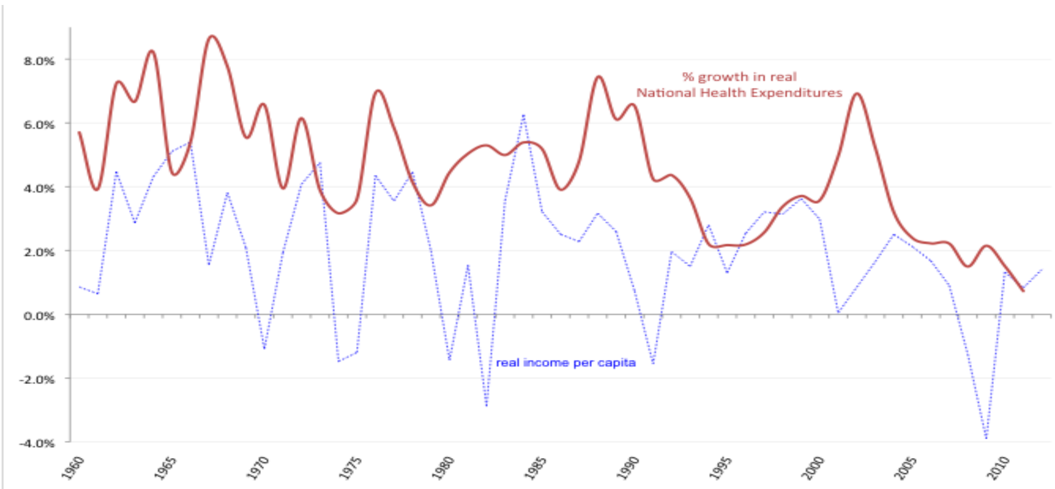
Table 1 Growth in real health expenditures (%) U.S., 1960–2009, lagged regression.

	Constant	real per capita GDP growth						Deflator		Time	**R2**
		lag 0	lag 1	lag 2	lag 3	lag 4	lag 5	lag 0	lag1		
Total NHE	.046	**.17**	**.07**	**.04**	**.19**	**.29**	**.23**	-.28	-.12	-.0006	**.702**
Hospital	.068	**-.17**	**.06**	**.06**	**.14**	**.39**	**.24**	-.15	.25	-.0011	**.705**
Physician	.044	**.03**	**.36**	**.14**	**.13**	**.37**	**.03**	-.52	-.69	-.0006	**.312**
Dental	.009	**.36**	**.16**	**.22**	**.18**	**.22**	**.32**	-.41	.07	-.0002	**.311**
Pharmaceutical	-.071	**.73**	**.34**	**.70**	**.71**	**.09**	**.15**	-1.04	-.95	.0016	**.457**
Long-term care	.058	**.22**	**.65**	**.45**	**.18**	**.54**	**.35**	-.67	.03	-.0013	**.662**
Insurance admin	.064	**.17**	**.44**	**.26**	**.34**	**.57**	**.49**	-.57	.12	-.0013	**.470**
Out of pocket	-.006	**.43**	**.26**	**.13**	**.11**	**.26**	**.00**	-.57	-.53	-.0001	**.245**

Source: www.cms.gov/nationalhealthexpenddata, accessed May 22, 2011

National expenditures and income are proportional to population, and to inflation. Such proportionality generates a superficial and economically meaningless correlation unless the analysis takes account of these two factors. Once adjusted, a graph of real (deflated) per capita national health spending and income (**figure 5a**) shows health spending has grown more rapidly on average (4.5 percent vs. 2.0 percent) but is trending down (from 7 percent to 2 percent).

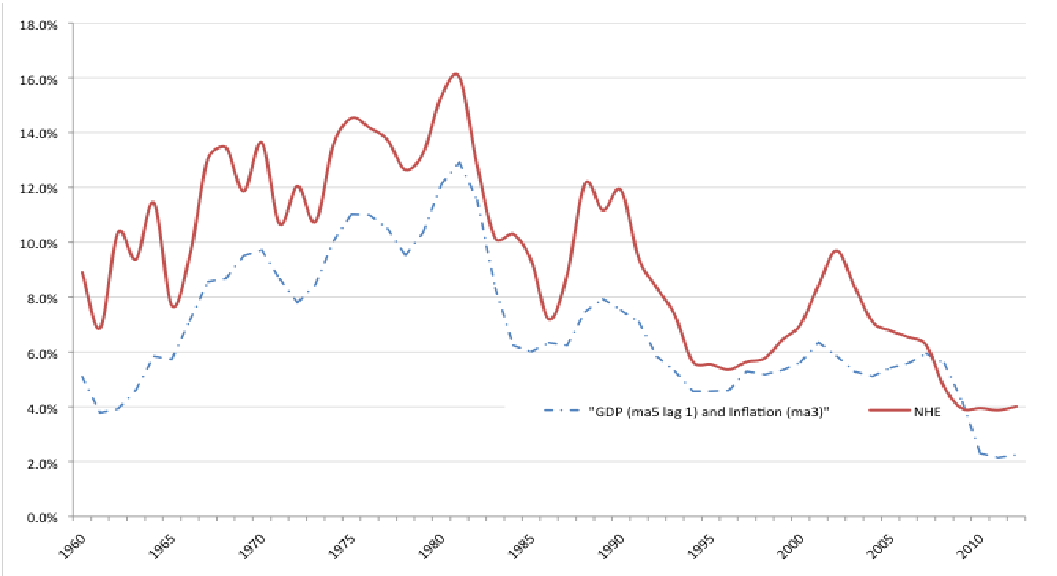
Figure 5a real per capita income and health spending



Source: CMS Office of the Actuary, National Health Expenditure Data.

The delayed response of health spending to changes in GDP and inflation is not so readily apparent in **figure 5a**, but is evident in ordinary least squares (OLS) (**table 1**) or autoregressive integrated moving average (ARIMA) time-series analyses. Most of the business cycle effects can be made visible by plotting “smoothed” real per capita income (six-year moving average [M.A.] for GDP, three-year M.A. for deflator) as a baseline for health expenditures (**figure 5b**).

Figure 5b real per capita national health expenditures and smoothed income



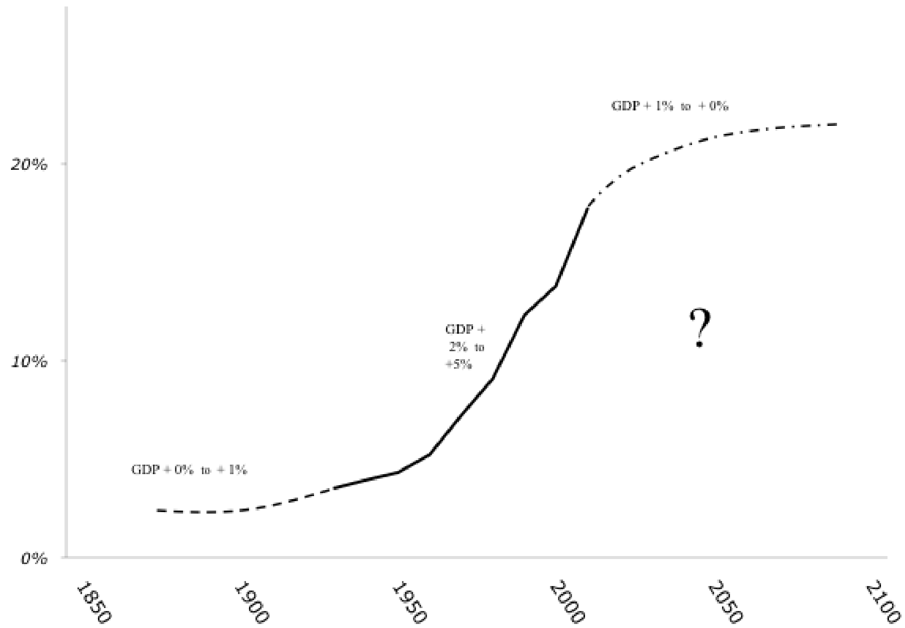
Source: Author calculations and smoothing of CMS Office of the Actuary, National Health Expenditure Data.

The annual percentage growth rate of per capita income, although variable (standard deviation [s.d.] = 2.1 percent), appears to be stationary and integrated of order 1. Indeed, despite two world wars and a great depression, real per capita growth in each 20-year period throughout the entire 20th century stayed close to 2 percent. Health spending is not integrated to the same degree but trends downward, and thus would appear to require further differencing to become stationary. This poses a number of econometric problems. Is it appropriate to measure variability relative to a simple mean (4.5 percent), or should it be relative to a trendline stretching down from 7 percent to 2 percent? How many times should the health series be differenced to become stationary for an ARIMA analysis? Can the two series be treated as co-integrated? Decomposition clarifies and resolves a number of these problems. It indicates that excess growth has slowed dramatically from 1970 to 1990. However, that slowing itself poses more questions—questions that cannot be answered until the span of observation is expanded beyond 50 years.

3.3 The Long Run

The rapid growth of health expenditures must ultimately slow down, if for no other reason than that health care (and other activities) cannot consume more than 100 percent of the total budget. The Congressional Budget Office (CBO), Centers for Medicare & Medicaid Services (CMS) Office of the Actuary and the Society of Actuaries (SOA)/Getzen models all project convergence toward a steady state at GDP + 0 percent beyond some future year (i.e., a constant share of GDP). Each, however, projects a slightly different year and stable share. Examining U.S. national health accounts for 1960–2012 (and for sporadic years previously), along with the CBO, CMS and SOA projections, supplemented by counts of physicians and other health care occupations and early 19th century estimates of consumption by industrial workers, provides support for the following generalization: Health care spending in the United States grew slowly for many decades up until almost 1960, soared rapidly until the 1980s and then began to moderate, slowing toward eventual stability at some point in the future. This would imply something like a standard S-shaped logistic growth curve as illustrated in **figure 6**.

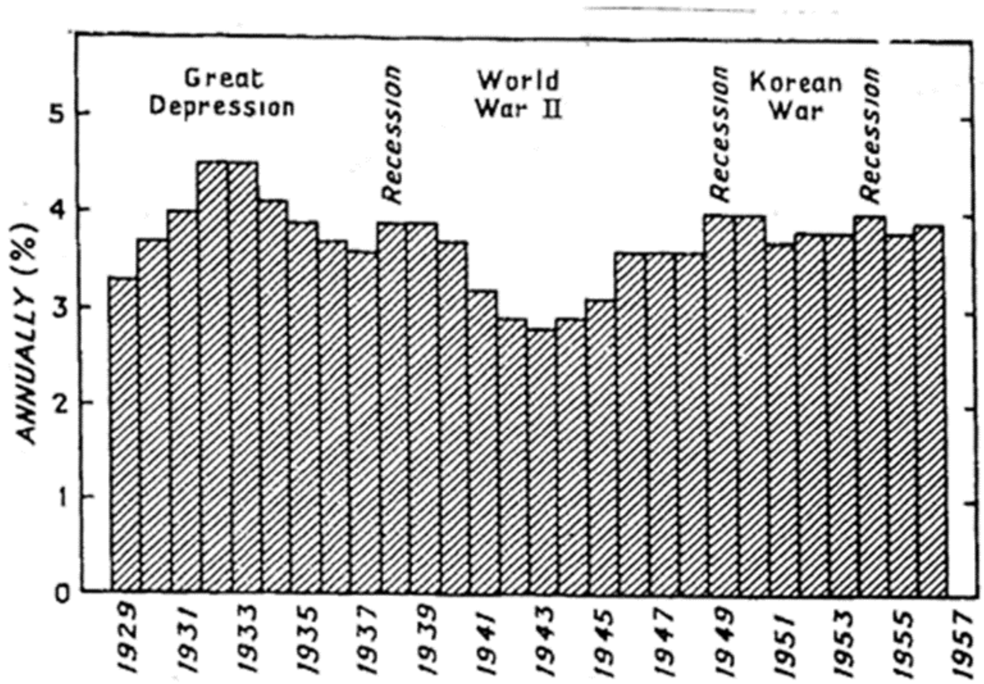
Figure 6 S-shaped logistic growth curve for health care of GDP projected from 1850 to 2100



Source: Author calculations and smoothing of CMS Office of the Actuary, National Health Expenditure Data.

The CBO, CMS and SOA projections are only projections, and the quality of estimates prior to 1960 is less refined and robust than the official U.S. and Organisation of Economic Co-operation and Development (OECD) national health accounts, yet an essentially S-shaped logistic growth curve for medical costs over the long run similar to **figure 6** is likely to apply for the United States and other developed countries with a modern health system, even though some details (and, of course, the future path) may be uncertain. The first published article assessing the trend in national health expenditures was written by J.R. Seale for *The Lancet* in 1959. He examined spending from 1929 through 1956, observed that spending remained near 4 percent of gross national product (GNP), fluctuating between 3.4 and 4.5 percent due mostly to business cycles without a discernable trend up or down. Seale concluded, “The proportion of the gross national product of a nation devoted to medical care tends to remain constant. It rises during national economic depressions and it falls during wars. A persistent rise in real per capita gross national product will tend to result in a very gradual increase in the proportion.” See **figure 7**.

Figure 7 “Stable” historical U.S. spending as percent of GDP 1929–57



Source: Reproduced from Seale (1959, figure 2)

Health spending had already begun to accelerate rapidly and make this conclusion outdated by the time Seale’s article appeared in 1959. By 1960, medical costs were up 53 percent over five years while GDP rose only 27 percent, pushing the health share up to .052. From 1960—65, health expenditure increases exceeded total economic growth by 2.5 percent annually. After the advent of Medicare in 1965, expenditures soared 79 percent over the next five years, exceeding GDP growth by 4.3 percent per year and bringing the health share to .072 in 1970. Medical spending continued to outpace GDP throughout the 1970s, 1980s and early 1990s. In 1994, health spending dipped below other sectors of the economy. For the next five years, expenditures were stable at

.138 of GDP. Excess growth resumed in 2001 but has averaged less than 1.5 percent over the last 10 years. Current national health spending hovers around .179 of GDP (Keehan et al. 2012).

The inflexion point of the S-curve where growth was most rapid appears to have come around 1982. At that time, excess growth was more than 5 percent per year, sufficient to have doubled the health share of GDP within 15 years if the rate of growth had not begun to slow down. With the passage of 30 years, it is now fairly evident that “bending the curve” toward eventual stabilization near GDP + 0 percent has, however intermittent, punctuated or jerky, started to occur. Spending is by no means under control, but the cumbersome political, social, financial and medical changes necessary to do so are slowly happening. The future is, of course, unpredictable. It may be that the bouts of moderation during the last several decades are a deceptive guide to what will happen by 2050 or 2102, yet a path looking something like a logistic S-curve is more likely than continued rapid growth that pushes medical expenditures above 40 or 50 percent of GDP.

Changing the span and unit of observation changes the pattern of results. Growth appears to be a rather smooth straight line over the course of a year or two, with much less variability than total economic activity (GDP or employment) (**figures 1** and **4a**). During a span of one or two decades, GDP and employment will usually have one or two sharp falls due to a recession. This recession will show up in health spending as a delayed shallow dip spread out over the following 36 to 72 months (**figure 4b**). Health spending will also experience some minor variation in yearly growth due to policy, accidents or random unspecified factors, yet the path of health spending is still much smoother than the path of GDP. The relative variability is reversed over the long run, with GDP growth appearing smoother and steadier than health care when viewed from a greater distance. Throughout the last century, real per capita income growth has averaged about 2 percent (s.d. = 0.8 percent) over each 20-year span. In contrast, the health sector soared after 1950 to quickly double, then triple its share of GDP before suddenly slowing from 1994 to 1999, and then again from 2009 to the present as it bent back to match the slower and steadier growth of GDP.

3.4 Trends

If the span of observation is too long, or too short, the phenomena are not readily apparent in first differences. They blend into the average, or become part of a slowly evolving trend. For example, a rainstorm may cool off a hot summer day. Observed for 600 seconds, there is little change—rain has merely raised or lowered the average temperature of that 10-minute span. Observed over five years, that rainstorm was just one inconsequential contributor to the month of April 2006. The cooling effect of that rainstorm will only be captured by using hourly measurements (or other unit of similar temporal length).

The ups and downs of a business cycle usually run four to eight years, but some potentially important determinants of health care spending (physician careers, hospital buildings, invention of a new treatment, drug patents, regulatory policy, population aging) may endure for many decades. If observed over a much shorter period (month, quarter, year), the longer-lasting effects of such changes will not show up as a peak or a trough, but only as a subtle change in the trend. All of the long-term forces, whether up, down or cyclical, blend together into a single "secular trend." The separate effects of aging, or technology, or physician practice are not readily observable. In regressions on annual rates of change such as those in **table 1** above, any such long-lasting effects create minor additions or subtractions to the constant but are otherwise invisible. Thus the effects of population aging on health spending may not be discernable in a single time-series. However, a comparison of time series in two countries, or a comparison across many countries at a single point in time, can allow an analyst to measure the effect of population aging on national health costs, as distinct from the effect of individual aging on the amount of total medical resources allocated to an elderly person.

3.5 Aging

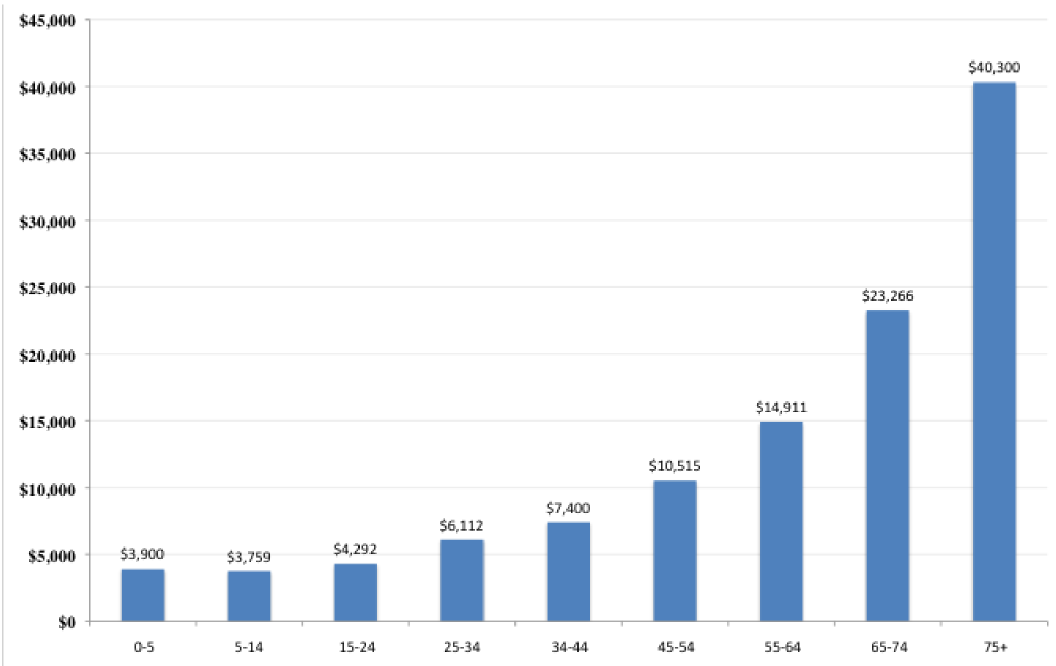
The contrast between population aging and individual aging illustrates the differences of phenomena at different organizational (micro and macro) and temporal (short and long) scales. It is well known that medical spending increases with a person's age. However, just because each individual's medical expenses tend to increase as each year passes does not imply that the medical expenses of the population as a whole increase with each passing year. Most observers see the system as a whole, with new individuals (births) replacing older ones (deaths) each year, so that the population average stays the same even as cost for each individual is steadily increasing. Seeing the system as a whole requires conscious effort. Doctors and their patients, who tend to age together, must ignore or reframe their daily experiences to see the bigger picture.

The phenomenon of population aging, the "graying of America," gathered a great deal of attention during the 1960s and 1970s. Life expectancy and medical costs were both growing rapidly, and it seemed natural to assume the relationship was causal. "Demographic" models were constructed that projected future health expenditures with a linear matrix which mimicked the format used for projections of future pension payouts (the "*c*" are "age-sex" categories or "age-sex-disease/disability" categories if more detail is desired).

$$\text{Total cost} = \sum cost_{\boldsymbol{c}} \times pop_{\boldsymbol{c}} \times (\text{total population growth} \times \text{other cost growth})$$

The fact the most sudden and dramatic cost increases were for diseases of the elderly made demographic explanations of medical cost growth seem compelling. Yet empirical investigations quickly showed change in the percentage of population age 65+ or number of elderly accounted for only a small portion of total cost increases, with most of the rise in national health expenditures attributable to increased cost per person (holding age and sex constant). One source of the illusory "age wave" can be seen in **figure 8**.

Figure 8 Spending by age category at a point in time



Source: Yamamoto (2013) and CMS Office of the Actuary, National Health Expenditure Data, Age and Gender Tables accessed April 2014.

A much different pattern of spending would occur without the extensive coverage available today. Spending by the elderly on medical care in 1950s was above average (150 percent) but still heavily constrained by personal and family budgets (**table 2**). The advent of Medicare and the expansion of employer insurance removed most individual budgetary constraints and old-age spending soared, eventually exceeding the average amount spent by those under age 65 by a ratio of more than 5:1. Note, however, that since 1990, the relative rates of growth in spending for the elderly have been below average and moved closer to the mean.

Table 2 Spending per person by age category*

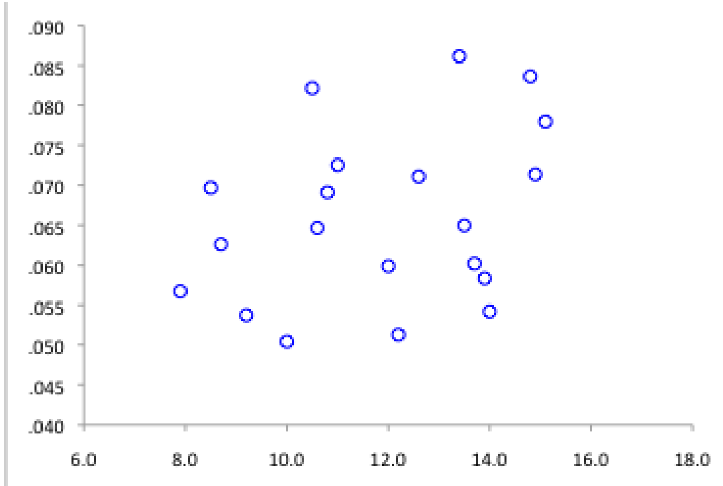
	1953	1963	1970	1977	1987	1996	2000	2004
All individuals	$69	143	291	616	1,664	3,153	3,803	5,276
Under age 65	$65	127	234	452	1,088	2,115	2,650	3,953

Age 65+	$109	299	809	1,962	5,830	10,285	11,778	14,797
**in 2009 nominal dollars*								
Ratio: over/under	**1.7**	**2.4**	**3.5**	**4.3**	**5.4**	**4.9**	**4.4**	**3.7**

Sources: Cutler and Meara (1997); Meara and Cutler (2004); Hartman (2008); Getzen (2013a)

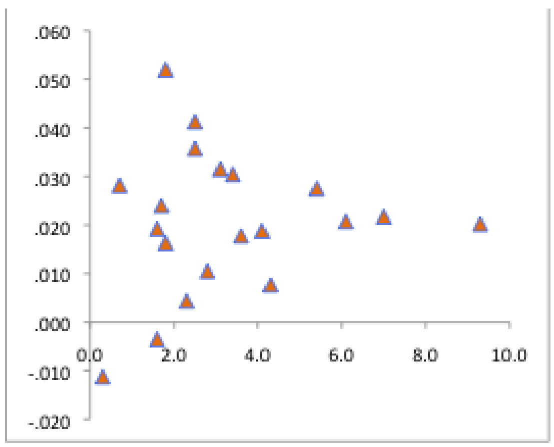
The conclusion that the correlation between population aging and medical cost growth was spurious is even more evident from a macro perspective. The process of economic development leads to both higher incomes and better health, so that a superficial examination would indicate a strong correlation between per capita medical spending and life expectancy. The undeveloped poor countries cluster in the lower left quadrant close to the origin (low income, low life expectancy) while the economically developed nations are clustered in the upper right (higher income, longer life expectancy). However, the share of GDP spent for health and percent of population over age 65 among OECD countries (**figure 9a**) shows essentially zero correlation. Using "share of GDP" rather than "per capita spending" normalizes the data to adjust for exchange rates and the unitary elasticity of incomes and wages, removing several sources of spurious correlation. **Figure 9a** is cross-sectional analysis, using macro observations, showing older nations do not have higher health spending at a point in time (e.g., United Kingdom). **Figure 9b** is a very basic panel analysis (cross-section of time series) showing countries whose populations are most rapidly aging are not those whose health share (or per capita spending) is rising most rapidly (e.g., Japan).

Figure 9a Correlation of health share of GDP with percent of population age 65+ (1975) is small.



Source: OECD Health Data.

Figure 9b Correlation of change in health share of GDP with change in percent of population 65+, 1975–2000



Source: OECD Health Data.

The idea that "aging" causes health care costs to rise has been incredibly persistent despite the mounting pile of contrary empirical evidence because it so obviously holds true for each person and thus makes "common sense" that it must hold true for the nation. This atomistic fallacy that what is true for the individual must be true for the group was reinforced by the fact that the increase in longevity and the increase in health care costs occurred at the same time in the development of most modern economies. The lack of empirical evidence for time-varying correlation between national aging and increased national health expenditures reinforces the hypothesis that they are co-determined. The two aspects of economic development tend to move together, but one does not cause the other, or vice versa, in any simplistic way.

More developed countries have more wealth and old people and spend more on medical care. These countries also have higher rates of urbanization, literacy, automobile ownership and financial services, and lower rates of fertility, infant mortality, civil unrest and transactions costs. The direct causal connection, to the extent one exists, is the ordinary economic one between income and spending—when the budget becomes larger, expenditures inevitably increase. However, even though aging per se is not a major cause of the rise in average medical costs, longevity is costly, and medical costs are a factor in that increasing financial burden.

3.6 Longevity is Costly

In 1900, with average medical expenses at 3 percent of GDP and average life expectancy of 47 years, the majority of workers did not have, or need, a financial plan for long-term care, retirement or health insurance.[3] Longevity has made all three a financing problem, one that exceeds the capability of the typical individual or family, and a burden increasingly shunned or avoided by employers and, to some extent, even by family members. A common financial feature of this trio is that they are not elements of the daily consumption of working families, so that each requires prefunding through savings or external transfer payments.

Medical care is fundamentally different from most markets in that those who pay (people earning money) are not those who benefit (people who are sick). In the formative era of employer

[3] The most common forms of health insurance sold at the turn of the century were actually for wage loss and funeral expenses rather than medical bills.

health insurance in the United States following World War II, the membership overlap between these two groups was substantial—health insurance was mostly traditional group risk pooling, with many of those paying premiums reasonably expecting some benefit for themselves or their families. Those whose health or work situation automatically excluded them (the infirm, elderly, poor children, those with chronic disabilities) were gradually brought under the financial umbrella through tax-funded social insurance plans (Medicare, Medicaid), and special arrangements were made where life-saving medicine was available but too expensive to cover, such as with end stage renal disease (ESRD). This mix of insurance plans was fine in the 1950s with medical costs under 5 percent of GDP and less than 10 percent of the population over age 65. The burden of medical care for the elderly was about 1 percent of income, a fraction that could relatively easily be funded through taxes or savings. By 2010, medical financing was in crisis and falling apart. The fact that the population over age 65 had risen to 13 percent contributed, but the real problem was that medical expenditures had tripled, to above 17 percent of GDP, and that relative medical costs per person for the elderly had quadrupled from 66 percent to 275 percent. Medical care for the elderly had become five times as costly, more than 5 percent of total income (GDP). The failures of the existing patchwork system of public and private coverage portend to be even worse by 2020, when the costs of medical care are projected to reach 20 percent of GDP, with the elderly accounting for almost half of that.[4]

Longevity makes health insurance ever more necessary. With advancing age, the overlap between payers and patients attenuates as it becomes less and less common for those currently working to receive benefits. The distinction between the costs of medical treatment, long-term care and retirement savings seems clear when applied to workers in their prime. The categories of

[4] Calculation of elderly fraction of NHE = (percent 65+) x (relative cost)/[(percent <65) + (percent 65+) x (relative cost)]

mobility, transportation, household assistance, health care and ordinary living expenses are much more blurred at age 70 and may be almost indistinguishable by age 90.

The question of whether or not something is a “medical” expense usually arises with regard to health insurance. Whether it is a pill, a diagnostic exam, a walker or a surgical procedure, the answer usually hinges on an administrative interpretation of a densely worded contract, not the consumer’s assessment or understanding of what is medical or necessary. The most common response to this anxious inquiry is “submit it to the carrier, and we will see if it is covered,” not “ask your doctor.” The terminology that sets the boundaries for determination of what is, or is not, a medical expense, is sometimes vague and far from immutable. The impact is straightforward—who pays. Medicare originally had a strict interpretation of “medical” that required “curative treatment” and excluded “care for maintenance of function or management of disability,” largely excluding hospice, home health and nursing home bills from payment. Over the intervening decades, various new categories of long-term care (skilled nursing facilities, long-term acute care, rehab) were included, and in the 1990s home health was the most rapidly growing category of Medicare costs.

Whether deemed medical care, long-term care or retirement assistance, the bundle of age-related expenses coming after age 65 is now sufficiently large to require setting aside one-third of earnings through taxes, savings, employer benefits or other prefunded plans. These expenses cannot readily be separated, but they are also not readily combined. It is common to believe most medical care should be available to all but that most retirement should be based on savings and personal income, not transfer payments. Separation of categories seemed clearer and the financial requirements much smaller in the 1950s. Continued economic development will not resolve these tensions. The widening cracks in the current financing of long-term care, medical care and pensions make it both necessary and likely that new contractual and organizational solutions will soon arise.

4 Modeling the Cost of Medical Care for the Elderly

4.1 Extending the Model to Include Age Effects

The growth in spending for the elderly can be analyzed by including population aging and the ratio of costs between old and young to extend the forecast model of total medical expenditures described in the previous sections. A retrospective analysis of the available data is provided in **table 3** below. The first line presents previous results; the health care share of GDP quadrupled from 4 percent to 16 percent from 1953 to 2004. Over the same time span, the percentage of the population age 65+ rose from 8.5 percent to 12.4 percent, and the ratio of spending on old:young increased from 1.7:1 to 3.7:1. These two age-related factors pushed the fraction of total medical costs attributable to the elderly from 13 percent to 35 percent.

Table 3 Cost of medical care for the elderly

						Projected - - - -		- - - - -
	1953	1963	1987	2004	growth ‘53–‘04	2012	2025	2050
NHE share of GDP	**.041**	**.057**	**.110**	**.160**	2.7%	**.179**	**.205**	**.263**
spending ratio old:young	**1.7**	**2.4**	**5.4**	**3.7**	1.6%	**3.7**	**3.7**	**3.7**
(f) percent of pop 65+	8.5%	9.4%	12.2%	12.4%	0.7%	13.8%	18.8%	20.9%
percent of NHE for 65+	13%	20%	43%	35%	1.9%	37%	46%	50%
share of GDP for 65+	**.006**	**.011**	**.047**	**.055**	4.6%	**.067**	**.095**	**.131**
65+ medical costs*	$2.1	$6.9	$222	$657	**11.9%**	$1,062	$2,832	$13,202

* in billions

Source: Author’s calculations from Table 2 and Getzen (2013a).

As with health care spending in general, the main factor driving expenditures on care of the elderly is the growth of the overall economy. The 11.9 percent annualized rate of increase from 1953–2004 in spending for those age 65+ can be decomposed as growth of GDP, growth in the share of GDP devoted to medical care and growth in the percentage of health spending attributable to the elderly.[5] Just as the annualized rate of growth in GDP (7.0 percent) can be decomposed into components of real incomes per capita (2.1 percent), inflation (3.6 percent) and population (1.2 percent), the rise in the percentage of health care costs attributable to the elderly can be decomposed into growth in the fraction of population age 65+ (0.7 percent) and growth in the ratio of cost per elderly person relative to the mean (+1.3 percent).

Table 4 Annual rate of growth in health spending age 65+, 1953–2004
(decomposed into component factors)

	2.1%	Real income per capita	
	3.6	Inflation	
GDP	1.2	Population	**(7.0%)**
Medical	2.7	Health share ("excess")	**(9.9%)**
	0.7	Percent of population 65+	
65+ medical	1.2	Cost ratio old:average	**(11.9%)**

Source: Author's calculations from Tables 2 & 3.

The rise in cost of health care for the elderly is notable, but just as notable (and important for predicting trends) is that the rate of increase was more than twice as rapid during the first half of this period than the second half (14.7 percent vs. 6.6 percent). Also to be noted are the factors responsible for this difference. The aging of the population slowed considerably, but the main

[5] Four technical points are worth noting.

- The percentage rates of growth combine multiplicatively rather than additively.
- The age-cost ratio estimates rely on data for personal health expenditures, which typically comprise about 85 percent of total national health expenditures.
- The percentage rate of growth in health share of GDP is identical to the rate of "excess" cost growth above GDP as defined by CBO and CMS (i.e., %H = %GDP +%excess = %GDP + %Share).
- The cost ratio used to calculate the percentage of total health costs attributable to the elderly is relative to the mean (R_{OM}) rather than relative to the young (R_{OY}), and depends on the population age distribution as well as the relative costs of older and younger individuals. More formally, let f = fraction of population age 65+ and normalizing the cost of younger persons at 1.0, then
    - average cost per person is (mean) = (1 − f)*1 + f *(ROY) hence ROM = ROY/[1 + f *(R_{OY} – 1)]
    - and the percent of total costs attributable to the elderly is f*[ROY/(1 + f *(R_{OY} – 1)].

factor causing the change in trend was a decline in the cost ratio, which was 1.68 in 1953, rose rapidly to 5.36 by 1987 and then fell to 3.74 in 2004. This, the distribution of resources between old and young, is malleable and open to policy directives, and is a major driver determining how large the financial burden of caring for the elderly will be in 2050. The only other factor that comes close to being this significant is the rise in average health care costs—that is, the distribution of resources between health and other demands. It is perhaps not purely coincidental that excess growth in health spending was somewhat greater before 1987 than after (2.9 percent vs. 2.7 percent) as reducing the extra spending on the elderly (relative to the average patient with comparable illness) is one element of a policy to control overall medical costs.

4.2 Projections and Uncertainties

If the cost ratio old:young remains 3.75 as it was when last estimated in 2004, then the percentage of health expenditures attributable to age 65+ will rise to 50 percent by 2050. Accepting the current CMS estimate that the health share of GDP will be 19.2 percent of GDP in 2020 and grow about 1 percent each year after that, reaching 26 percent by 2050 implies that the spending on the elderly will almost double by 2020 and grow more than tenfold by 2050 to exceed $1.3 trillion. The relative share accounted for by the elderly will rise more and more slowly after 2050, as the bulge of the baby boom passes and resistance is created both by the policy and the math of the distribution.

Assuming the cost ratio old:young stays at 3.7 is a large and uncertain "if." In the previous five decades, it ranged from a low of 1.7 to a high of 5.4. The current assumption placing the age 65+ share of health spending at 50 percent in 2050 would move down to 40 percent if the cost ratio were 2.5 and lift to 54 percent if the ratio were 4.5. Using the top and bottom of the range would imply even larger uncertainty. Although future cost ratios that depend on the vagaries of a complex health system and the whims of legislators can be expected to move somewhat unpredictably within a sizable range, the essential demographic factors (population growth and fraction of the population age 65+) are much less uncertain and lie within a much smaller range, even out to 2050 and beyond. The current Census Bureau projections for *f*, the fraction of the population age 65+, have a 2060 low of 21.3 percent and high of 22.6 percent, a range of ±0.7 percent (U.S. Census Bureau 2013).

The rate of increase in national health expenditures in excess of GDP is likely to be the second largest source of uncertainty. The CMS estimate for excess growth averages 1.5 percent for the last five years of the projection and is used as a default baseline here. However, anything from 0.5 percent to 2.5 percent can be quite reasonably defended. A quantification of the range of uncertainty is speculative at this point, but the likelihood of being inside that ±1 percent range is probably on the order of 90 percent (Getzen 2013a). Uncertainty regarding the rate of growth in GDP is perhaps even larger, and certainly more important in terms of public welfare, but falls outside the scope of health care forecasting. The CMS projected rate of long-run growth in real income per capita of 1.4 percent is reasonable as an estimate of central tendency, but the average actual rate exceeded 2.0 percent for the last half of the 20th century, and the recent recession has so shaken the confidence of some economists that they predict long-run average growth of just 1 percent or less (Gordon 2010, 2012). Inflation, assumed by CMS to lie mostly between 1 percent and 4 percent, is generally considered to be almost unpredictable over the long run—and also essentially irrelevant as the extent to which prices, costs, wages, taxes, etc., all move together.

This forecast of future health spending for the elderly depends heavily on two distributional parameters: what share of total resources available should be spent on health (health share of GDP) and what fraction of that health spending should be devoted to the elderly (percent of health spending for 65+). Both are determined primarily by politics and social choice rather than demographics or biology. The major source of uncertainty in the forecast is the complexity and uncertainty of shared decisions, not the uncertainty of mortality or the biological complexity of diseases. The major demographic factors are mostly immune to policy, and even out to 2060 are reasonably predictable. The purely macro-economic factors (GDP, inflation) are more uncertain, and can perhaps be influenced by policy, but certainly fall outside the remit of a health care forecast.

4.3 Micro or Macro: Why Not Use Detailed Demographic and Biologic Categories?

Most projections for the cost of health care to the elderly begin at the individual level, creating detailed weights by age, sex and morbidity category; multiply each by a category-specific disease incidence rate, then by a cost per illness episode, before finally summing costs across all categories and types of illnesses as in the equation below.

$$\text{Total cost} = \sum(\textit{age, sex, morbidity category}) \times (\textit{incidence rates}) \times (\textit{cost per case})$$

Such models often encompass hundreds or even thousands of computation, although it has become common to try to estimate the cost just for one disease (diabetes, stroke, AIDS), type of patient (hospice, obese, blood pressure >140) or provider (ambulatory surgery centers, emergency rooms, MRIs). The main empirical weakness of relying on such detailed category decompositions for long-run projections is that the current detailed estimates for hundreds of weights, rates and itemized costs must still be extrapolated into the future using a multiplier for expected average increase per year—a multiplier that is usually much larger, and more uncertain, than the measured effects of changes in the mix of weights, rates and costs. The essential conceptual weakness is the focus on the individuals and the components, rather than the system. In particular, such categorical extrapolations tend to ignore a central fact: Spending is above all a budgetary decision, with a total budget constraint imposed at the national level, rather than the outcome of individual events or decisions subject only to individual budget constraints.

The macro forecast model used here has three elements: the amount of money spent each year, the share of that spent on health and the fraction of the health share devoted to the elderly. This is an extreme simplification, an abstraction with considerable distance from ordinary medical reality. It forces the analyst to concentrate on the system as a whole rather than the individual parts. It abstracts away from the many fascinating details of doctors, patients, microbes, oncogenes and individual economic decisions and payments. Such simplification might not be worthwhile if it did not lead to a considerable improvement in accuracy—which it does, routinely yielding far more accurate and comprehensible results than the many intricate large-scale demographic projections of cost by disease category. The ordinary fluctuations due to business cycles and the long-run super-cycle marking the development and maturation of the modern medical system become much clearer with a macro model. Correctly done, abstraction and simplification also make the implications for policy much more apparent and provide ways to measure the progress (or lack thereof) with regard to the most important questions.

Summary of the advantages of macro health modeling:

- Accuracy is better (especially when forecasting rather than backcasting)
- Empirically sounder, incorporating the central budget constraint
- Concentrates on largest sources of uncertainty
- Focuses attention on the system, not the parts
- Highlights policy relevant variables rather than technical details or immutable facts

- Clarifies the essential choices: What share of GDP? What percent for the elderly?
- Simplification allows time for thought, analysis of long-run determinants and disturbances

5 Discussion: Financing Medical Care, LTC and Retirement

The growth of health spending looks different when examined by the month, the year or the decade. The data reveal different phenomena, or different aspects of an evolving phenomenon, as the scale and unit of observation is changed. The main conclusion reached from a short span is that growth in the health sector is much more stable and continuous than for the rest of the economy. The "curve" looks like a straight line. Examined over 10 to 20 years, the health sector still appears to be relatively steady but does show some evidence of moving up and down in a delayed and smoothed reflection of the business cycle. Only from a very long-run perspective across many decades does it become apparent that health spending has gone through a tremendous growth spurt. Instead of a steadily rising linear trend, it has the typical S-curve of a logistic growth process.

Longevity also increased rapidly in the United States during the 20th century, yet population aging was not the cause, or at least not a primary or proximate cause, of the rise in health care spending. Quantitatively, the extra cost of treating each disease, and the extra willingness to treat diseases at advances ages, were far more important than increases in the average age or fraction of the population that were elderly. Instead of trying to discern causality, it is probably more useful to see "aging" and "advanced expensive medical care" as integral parts of the larger social process that constitutes modern economic development, akin to transportation, urbanization, leisure time, professional sports, literacy, universities, manufacturing, finance, delayed marriage, decline of farming, corporations, communications, risk management and so on.

The United Kingdom, Europe, Japan and the United States went through the development process during the 19th and 20th centuries; China, India, Brazil and Turkey are undergoing similar transformations in the 20th and 21st, although perhaps at an accelerated pace. All of them have or will experience increasing longevity and rising medical costs, with both factors making pooled financing ever more necessary, although not necessarily more easy to legislate and implement.

Building financial institutions for medical care (and for retirement and LTC) are part of a process of economic and social development. The trio of expenses related to longevity that must be prefunded through savings or taxes will realistically take 1/3 or more of earnings during the 21st century. Some aspects and variables are fairly predictable. Demographic statistics are among the best and most predictable, making the age-sex distribution of the elderly population in 2050 fairly certain. Despite the clarity and confidence of demographic projections, some of the mechanisms and causality are still unclear (fertility declines, or lack thereof; ebb and flow of immigration) and could create major shifts in the future.

The long-run growth rate of GDP will have the greatest impact on future medical costs, closely followed by technology. Per capita GDP growth has averaged 2.2 percent over the last 100 years, but following a disastrous start to the new millennium, many commentators expect slower growth in the future (Gordon 2010). Advances in medical science are neither predictable nor quantifiable, making projections of GDP seem comparatively sound and well understood. Even so, the difference between income growth of 1.5 percent and 2.5 percent annually compounded over 75 years amounts to more than 300 percent of current resources for each person in real terms.

Policy cannot do much to affect demography or GDP growth or the rate of new medical discoveries.[6] It can change the financing mechanisms for medical care, long-term care and

[6] More precisely, policy can disrupt and retard some processes, but cannot do much to move GDP or technology forward faster in more favorable directions.

retirement. Any attempt to do so will be contentious because it would necessarily challenge the values of some portion of the public and change the distribution of resources across families. Despite rhetoric to the contrary, such a debate must rely on politics rather than science or economic efficiency. The current problems of financing longevity are contingent upon some core unresolved questions. Does everyone deserve a basic level of medical care? What fraction of costs should be borne through social insurance, and how much should remain private? There are also more pragmatic process questions, such as whether retirement, long-term care and medical care should each be financed separately, or have financing that treats them as an interrelated bundle of services. The likelihood that average life expectancy of the elderly will exceed 90 years is high, as is the likelihood that per capita income will exceed $100,000. Finding a satisfactory combination of savings, subsidies, social and private insurance to fund retirement, long-term care and medicine for most or all Americans is rather less likely.

6 References

Abel-Smith, Brian. 1967. “An International Study of Health Expenditure and its Relevance for Health Planning.” Public Health Papers No. 32, World Health Organization, Geneva.

Arrow, Kenneth J. 1963. “Uncertainty and the Welfare Economics of Medical Care.” *American Economic Review* 53(5): 141-149.

Commonwealth of Massachusetts. 1875. *Sixth Annual Report of Bureau of the Statistics of Labor*. Public Document No. 31. Boston: Wright & Potter, State Printers.

Cutler, David and Ellen Meara. 1997. “The medical costs of the young and old: A forty-year perspective.” NBER working paper #6114 accessed at www.NBER.org.

Cylus, Jonathan et al., 2011. “Pronounced Gender and Age Differences are Evident in Personal Health Care Spending Per Person.” *Health Affairs* 30(1):153-160.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Getzen, Thomas E. 1992. “Population Aging and the Growth of Health Expenditures.” *Journal of Gerontology: Social Sciences* 47 (3): S98–104.

———. 2000a. “Health Care is an Individual Necessity and a National Luxury: Applying Multilevel Decision Models to Health Care Expenditures.” *Journal of Health Economics* 19 (2): 259–70.

———. 2000b. “Forecasting Health Expenditures: Short, Medium and Long (Long) Term.” *Journal of Health Care Finance* 26 (3): 56–72.

———. 2006. “Aggregation and the Measurement of Health Care Costs.” *Health Services Research* 41 (5): 1938–54.

———. 2007. “Modeling Long Term Healthcare Cost Trends.” Society of Actuaries, Schaumburg, IL. http://www.SOA.og/research/ health/research-hlthcare-trends.aspx (December, 2007 with spreadsheet updated November, 2012).

———. 2013a. *Health Economics and Financing*. 5th edition. Hoboken: John Wiley & Sons.

———. 2013b. “Accuracy and Range of Actuarial Projections of Health Care Costs.” Working paper. doi:10.2139/ssrn.2213648.

———. 2014. “Macroeconomic Dynamics of Health: Lags and Variability in Mortality, Employment and Spending.” In *Encyclopedia of Health Economics*, vol. 2, editor-in-chief, Anthony J. Culyer. Oxford: Elsevier.

Gordon, Robert J. 2010. “Revisiting U.S. Productivity Growth Over the Past Century with a View of the Future.” NBER Working Paper No. 15834, National Bureau of Economic Research, Cambridge, MA.

———. 2012. “Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds.” NBER Working Paper No. 18315, National Bureau of Economic Research, Cambridge, MA.

Hartman, M et. al. 2008. Heffler S. US health spending by age, selected years through 2004. *Health Affairs* 27(1):w1–12.

Keehan, Sean P., Gigi A. Cuckler, Andrea M. Sisko, Andrew J. Madison, Sheila D. Smith, Joseph M. Lizonitz, John A. Poisal, and Christian J. Wolfe. 2012. "National Health Expenditure Projections: Modest Annual Growth Until Coverage Expands and Economic Growth Accelerates." *Health Affairs* 31 (7): 1600–12.

Lassman, David et. al. 2014. "US Health Spending Trends By Age And Gender: Selected Years 2002–2010." *Health Affairs* 33(5): 815-822.

Meara, Ellen, Chapin White, and David M. Cutler. 2004. "Trends in Medical Spending by Age, 1963–2000." Health Affairs 23 (4): 176–83.

National Bureau of Economic Research. 2012. "U.S. Business Cycle Expansions and Contractions." http://www.nber.org/cycles/US_Business_Cycle_Expansions_and_Contractions_20120423.pdf.

Newhouse, Joseph P. 1977. "Medical-Care Expenditure: A Cross-National Survey." *Journal of Human Resources* 12 (1): 115–25.

Seale, J. R. 1959. "A General Theory of National Expenditure on Medical Care." *The Lancet* 274 (7102): 555–59.

Stigler, George J. 1954. "The Early History of Empirical Studies of Consumption Behavior." *The Journal of Political Economy* 42 (2): 95–113.

U.S. Bureau of Labor Statistics. 2013. "Consumer Expenditures in 2011." *BLS Reports* 1042 (April).

U.S. Census Bureau. 2013. "Population and Households: Age and Sex." Accessed September 15. www.census.gov/population/age.

Yamamoto, Dale H. 2013. "Health Care Costs—From Birth to Death." Independent Report Series Report 2013-1. Health Care Cost Institute, Washington, D.C.